

0300

#9



OIPE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/917,383

DATE: 10/01/2002

TIME: 16:15:50

Input Set : A:\4017us01.app

Output Set: N:\CRF4\10012002\I917383.raw

3 <110> APPLICANT: DING, SHI-YOU  
 4 ADNEY, WILLIAM S.  
 5 VINZANT, TODD B.  
 6 DECKER, STEPHEN R.  
 7 HIMMEL, MICHAEL E.  
 9 <120> TITLE OF INVENTION: THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS  
 10 CELLULOLYTICUS  
 12 <130> FILE REFERENCE: 40170.6US01  
 C--> 14 <140> CURRENT APPLICATION NUMBER: US/09/917,383  
 15 <141> CURRENT FILING DATE: 2001-07-28  
 17 <160> NUMBER OF SEQ ID NOS: 14  
 19 <170> SOFTWARE: PatentIn Ver. 2.1  
 21 <210> SEQ ID NO: 1  
 22 <211> LENGTH: 1228  
 23 <212> TYPE: PRT  
 24 <213> ORGANISM: Artificial Sequence  
 26 <220> FEATURE:  
 27 <223> OTHER INFORMATION: Description of Artificial Sequence: Segment of  
 28 GuxA  
 30 <400> SEQUENCE: 1  
 31 Met Glu Arg Thr Gln Gln Ser Gly Arg Asn Cys Arg Tyr Gln Arg Gly  
 32 1 5 10 15  
 34 Thr Thr Arg Met Pro Ala Ile Ser Lys Arg Leu Arg Ala Gly Val Leu  
 35 20 25 30  
 37 Ala Gly Ala Val Ser Ile Ala Ala Ser Ile Val Pro Leu Ala Met Gln  
 38 35 40 45  
 40 His Pro Ala Ile Ala Ala Thr His Val Asp Asn Pro Tyr Ala Gly Ala  
 41 50 55 60  
 43 Thr Phe Phe Val Asn Pro Tyr Trp Ala Gln Glu Val Gln Ser Glu Ala  
 44 65 70 75 80  
 46 Ala Asn Gln Thr Asn Ala Thr Leu Ala Ala Lys Met Arg Val Val Ser  
 47 85 90 95  
 49 Thr Tyr Ser Thr Ala Val Trp Met Asp Arg Ile Ala Ala Ile Asn Gly  
 50 100 105 110  
 52 Val Asn Gly Gly Pro Gly Leu Thr Tyr Leu Asp Ala Ala Leu Ser  
 53 115 120 125  
 55 Gln Gln Gln Gly Thr Thr Pro Glu Val Ile Glu Ile Val Ile Tyr Asp  
 56 130 135 140  
 58 Leu Pro Gly Arg Asp Cys Ala Ala Leu Ala Ser Asn Gly Glu Leu Pro  
 59 145 150 155 160  
 61 Ala Thr Ala Ala Gly Leu Gln Thr Tyr Glu Thr Gln Tyr Ile Asp Pro  
 62 165 170 175  
 64 Ile Ala Ser Ile Leu Ser Asn Pro Lys Tyr Ser Ser Leu Arg Ile Val

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/917,383

DATE: 10/01/2002

TIME: 16:15:50

Input Set : A:\4017us01.app

Output Set: N:\CRF4\10012002\I917383.raw

```

65          180          185          190
67 Thr Ile Ile Glu Pro Asp Ser Leu Pro Asn Ala Val Thr Asn Met Ser
68          195          200          205
70 Ile Gln Ala Cys Ala Thr Ala Val Pro Tyr Tyr Glu Gln Gly Ile Glu
71          210          215          220
73 Tyr Ala Leu Thr Lys Leu His Ala Ile Pro Asn Val Tyr Ile Tyr Met
74 225          230          235          240
76 Asp Ala Ala His Ser Gly Trp Leu Gly Trp Pro Asn Asn Ala Ser Gly
77          245          250          255
79 Tyr Val Gln Glu Val Gln Lys Val Leu Asn Ala Ser Ile Gly Val Asn
80          260          265          270
82 Gly Ile Asp Gly Phe Val Thr Asn Thr Ala Asn Tyr Thr Pro Leu Lys
83          275          280          285
85 Glu Pro Phe Met Thr Ala Thr Gln Gln Val Gly Gly Gln Pro Val Glu
86          290          295          300
88 Ser Ala Asn Phe Tyr Gln Trp Asn Pro Asp Ile Asp Glu Ala Asp Tyr
89 305          310          315          320
91 Ala Val Asp Leu Tyr Ser Arg Leu Val Ala Ala Gly Phe Pro Ser Ser
92          325          330          335
94 Ile Gly Met Leu Ile Asp Thr Leu Arg Asn Gly Trp Gly Gly Pro Asn
95          340          345          350
97 Glu Pro Thr Gly Pro Ser Thr Ala Thr Asp Val Asn Thr Phe Val Asn
98          355          360          365
100 Gln Ser Lys Ile Asp Leu Arg Gln His Arg Gly Leu Trp Cys Asn Gln
101          370          375          380
103 Asn Gly Ala Gly Leu Gly Gln Pro Pro Gln Ala Ser Pro Thr Asp Phe
104 385          390          395          400
106 Pro Asn Ala His Leu Asp Ala Tyr Val Trp Ile Lys Pro Pro Gly Glu
107          405          410          415
109 Ser Asp Gly Thr Ser Ala Ala Ser Asp Pro Thr Thr Gly Lys Lys Ser
110          420          425          430
112 Asp Pro Met Cys Asp Pro Thr Tyr Thr Thr Ser Tyr Gly Val Leu Thr
113          435          440          445
115 Asn Ala Leu Pro Asn Ser Pro Ile Ala Gly Gln Trp Phe Pro Ala Gln
116          450          455          460
118 Phe Asp Gln Leu Val Ala Asn Ala Arg Pro Ala Val Pro Thr Ser Thr
119 465          470          475          480
121 Ser Ser Ser Pro Pro Pro Pro Pro Ser Pro Ser Ala Ser Pro Ser
122          485          490          495
124 Pro Ser Pro Ser Pro Ser Pro Ser Ser Pro Ser Pro Ser Pro Ser
125          500          505          510
127 Pro Ser Ser Ser Pro Ser Pro Ser Pro Ser Pro Ser Pro Ser Pro Ser
128          515          520          525
130 Ser Ser Pro Ser Pro Ser Pro Ser Ser Ser Pro Ser Pro Ser Pro Ser
131          530          535          540
133 Pro Ser Pro Ser Pro Ser Ser Ser Pro Ser Pro Ser Pro Ser Ser Ser
134 545          550          555          560
136 Pro Ser Pro Ser Pro Ser Pro Ser Pro Ser Pro Ser Ser Ser Pro Ser
137          565          570          575

```

## RAW SEQUENCE LISTING

DATE: 10/01/2002

PATENT APPLICATION: US/09/917,383

TIME: 16:15:50

Input Set : A:\4017us01.app

Output Set: N:\CRF4\10012002\I917383.raw

```

139 Pro Ser Pro Thr Ser Ser Pro Val Ser Gly Gly Leu Lys Val Gln Tyr
140      580      585      590
142 Lys Asn Asn Asp Ser Ala Pro Gly Asp Asn Gln Ile Lys Pro Gly Leu
143      595      600      605
145 Gln Leu Val Asn Thr Gly Ser Ser Val Asp Leu Ser Thr Val Thr
146      610      615      620
148 Val Arg Tyr Trp Phe Thr Arg Asp Gly Gly Ser Ser Thr Leu Val Tyr
149 625      630      635      640
151 Asn Cys Asp Trp Ala Ala Met Gly Cys Gly Asn Ile Arg Ala Ser Phe
152      645      650      655
154 Gly Ser Val Asn Pro Ala Thr Pro Thr Ala Asp Thr Tyr Leu Gln Leu
155      660      665      670
157 Ser Phe Thr Gly Gly Thr Leu Ala Ala Gly Gly Ser Thr Gly Glu Ile
158      675      680      685
160 Gln Asn Arg Val Asn Lys Ser Asp Trp Ser Asn Phe Thr Glu Thr Asn
161      690      695      700
163 Asp Tyr Ser Tyr Gly Thr Asn Thr Thr Phe Gln Asp Trp Thr Lys Val
164 705      710      715      720
166 Thr Val Tyr Val Asn Gly Val Leu Val Trp Gly Thr Glu Pro Ser Gly
167      725      730      735
169 Thr Ser Pro Ser Pro Thr Pro Ser Pro Ser Pro Ser Pro Ser
170      740      745      750
172 Pro Gly Gly Asp Val Thr Pro Pro Ser Val Pro Thr Gly Leu Val Val
173      755      760      765
175 Thr Gly Val Ser Gly Ser Ser Val Ser Leu Ala Trp Asn Ala Ser Thr
176      770      775      780
178 Asp Asn Val Gly Val Ala His Tyr Asn Val Tyr Arg Asn Gly Val Leu
179 785      790      795      800
181 Val Gly Gln Pro Thr Val Thr Ser Phe Thr Asp Thr Gly Leu Ala Ala
182      805      810      815
184 Gly Thr Ala Tyr Thr Tyr Thr Val Ala Ala Val Asp Ala Ala Gly Asn
185      820      825      830
187 Thr Ser Ala Pro Ser Thr Pro Val Thr Ala Thr Thr Thr Ser Pro Ser
188      835      840      845
190 Pro Ser Pro Thr Pro Thr Gly Thr Thr Val Thr Asp Cys Thr Pro Gly
191      850      855      860
193 Pro Asn Gln Asn Gly Val Thr Ser Val Gln Gly Asp Glu Tyr Arg Val
194 865      870      875      880
196 Gln Thr Asn Glu Trp Asn Ser Ser Ala Gln Gln Cys Leu Thr Ile Asn
197      885      890      895
199 Thr Ala Thr Gly Ala Trp Thr Val Ser Thr Ala Asn Phe Ser Gly Gly
200      900      905      910
202 Thr Gly Gly Ala Pro Ala Thr Tyr Pro Ser Ile Tyr Lys Gly Cys His
203      915      920      925
205 Trp Gly Asn Cys Thr Thr Lys Asn Val Gly Met Pro Ile Gln Ile Ser
206      930      935      940
208 Gln Ile Gly Ser Ala Val Thr Ser Trp Ser Thr Thr Gln Val Ser Ser
209 945      950      955      960
211 Gly Ala Tyr Asp Val Ala Tyr Asp Ile Trp Thr Asn Ser Thr Pro Thr

```

## RAW SEQUENCE LISTING

DATE: 10/01/2002

PATENT APPLICATION: US/09/917,383

TIME: 16:15:50

Input Set : A:\4017us01.app

Output Set: N:\CRF4\10012002\I917383.raw

```

212          965          970          975
214 Thr Thr Gly Gln Pro Asn Gly Thr Glu Ile Met Ile Trp Leu Asn Ser
215          980          985          990
217 Arg Gly Gly Val Gln Pro Phe Gly Ser Gln Thr Ala Thr Gly Val Thr
218          995          1000          1005
220 Val Ala Gly His Thr Trp Asn Val Trp Gln Gly Gln Gln Thr Ser Trp
221          1010          1015          1020
223 Lys Ile Ile Ser Tyr Val Leu Thr Pro Gly Ala Thr Ser Ile Ser Asn
224          1025          1030          1035          1040
226 Leu Asp Leu Lys Ala Ile Phe Ala Asp Ala Ala Ala Arg Gly Ser Leu
227          1045          1050          1055
229 Asn Thr Ser Asp Tyr Leu Leu Asp Val Glu Ala Gly Phe Glu Ile Trp
230          1060          1065          1070
232 Gln Gly Gly Gln Gly Leu Gly Ser Asn Ser Phe Ser Val Ser Val Thr
233          1075          1080          1085
235 Ser Gly Thr Ser Ser Pro Thr Pro Ser Pro Ser Pro Thr Pro Thr Pro
236          1090          1095          1100
238 Ser Pro Thr Pro Thr Pro Ser Pro Ser Pro Thr Pro Ser Pro Ser Pro
239          1105          1110          1115          1120
241 Thr Ser Ser Pro Ser Ser Ser Gly Val Ala Cys Arg Ala Thr Tyr Val
242          1125          1130          1135
244 Val Asn Ser Asp Trp Gly Ser Gly Phe Thr Ala Thr Val Thr Val Thr
245          1140          1145          1150
247 Asn Thr Gly Ser Arg Ala Thr Asn Gly Trp Thr Val Ala Trp Ser Phe
248          1155          1160          1165
250 Gly Gly Asn Gln Thr Val Thr Asn Tyr Trp Asn Thr Ala Leu Thr Gln
251          1170          1175          1180
253 Ser Gly Ala Ser Val Thr Ala Thr Asn Leu Ser Tyr Asn Asn Val Ile
254          1185          1190          1195          1200
256 Gln Pro Gly Gln Ser Thr Thr Phe Gly Phe Asn Gly Ser Tyr Ser Gly
257          1205          1210          1215
259 Thr Asn Ala Ala Pro Thr Leu Ser Cys Thr Ala Ser
260          1220          1225
263 <210> SEQ ID NO: 2
264 <211> LENGTH: 3687
265 <212> TYPE: DNA
266 <213> ORGANISM: Artificial Sequence
268 <220> FEATURE:
269 <223> OTHER INFORMATION: Description of Artificial Sequence: Segment of
270      GuxA
272 <400> SEQUENCE: 2
273 atggagcgaa cccaacaatc cggacggaac tgcaggtacc agagaggaac gacacgaatg 60
274 cccgccatct caaaacggct gcgagccggc gtctctgccg gggcggtgag catcgagcc 120
275 tccatcgtgc cgctggcgat gcagcaccct gccatcgccg cgacgcacgt cgacaatccc 180
276 tatgcgggag cgaccttctt cgtcaaccgc tactgggagc aagaagtaca gagcgaagcg 240
277 gcgaaccaga ccaatgccac tctcgcagcg aaaatgcgcg tcgtttccac atattcgacg 300
278 gccgtctgga tggaccgcat cgctgcgacg aacggcggtc acggcggacc cggcttgacg 360
279 acatatctgg acgccgccct ctcccagcag cagggaacca cccctgaagt cattgagatt 420
280 gtcactctacg atctgccggg acgcgactgc gcggcgctcg cctccaacgg cgaactgccc 480

```

## RAW SEQUENCE LISTING

DATE: 10/01/2002

PATENT APPLICATION: US/09/917,383

TIME: 16:15:50

Input Set : A:\4017us01.app

Output Set: N:\CRF4\10012002\I917383.raw

```

281 gctacggcag caggtttgca gacctatgaa acgcagtaca tcgatccgat tgcgagtatc 540
282 ctgagcaatc cgaagtactc cagcctgcgg atcgtgacga tcattgagcc ggactcgctg 600
283 ccaaacgcgg tcaccaatat gagcattcaa gcgtgtgcaa cggcgggtgcc gtattacgag 660
284 caaggcatcg agtacgcgct cagcaaattg cagccattc cgaacgtgta catctacatg 720
285 gacgccgccc actccggctg gcttgggtgg cccaataatg ccagcggata cgtacaggaa 780
286 gtccagaagg tcctcaacgc gagcatcggg gtcaacggca tcgacggctt cgtcaccaac 840
287 acggcgcaatt acacgccgtt gaaggagccg ttcatgaccg ccaccagca ggtcggcgga 900
288 cagccgggtg agtcggcgaa tttctaccag tggaatcctg acatcgacga agccgactac 960
289 gcggttgact tgtactcgcg gctcgtcgcc gctggctttc caagcagcat cggcatgctc 1020
290 atcgacacct tacgcaacgg ttgggggtgt ccgaacgaac caacaggccc gagcaccgag 1080
291 accgatgtca acaccttcgt caaccagtcg aagattgacc ttcggcagca ccgcgccctg 1140
292 tgggtgcaacc agaacgggtg gggcctcgcc cagccgccgc aggcaagccc gacggacttc 1200
293 ccgaacgcgc acctcgacgc gtatgtcttg atcaagccgc cgggtgagtc ggacggcaca 1260
294 agcgtcgcca gcgatccgac aactggcaag aagtcggacc ccatgtgcga cccgacgtac 1320
295 acgacgtcgt acggggtaact gaccaacgcg ttaccgaact ccccgatcgc cggccagtgg 1380
296 ttcccgcgcg agtttgacca gcttgtcgcg aacgcacggc cagcggtgcc gacgtcgacc 1440
297 agctcgagcc cgcgcctcc gccgcgagt ccgtcggctt cgccgagtcg gagcccagat 1500
298 ccgagcccga gcagctcgcc atcgccgctg ccgtctccga gctcgagccc gtctccgctg 1560
299 ccgagcccga gtccgagccc gagtagctcg ccgtcgccgt ctccgagctc gagcccgtct 1620
300 ccgtcgccga gcccgagtc gagcccagat agctcgccgt cgccgtctcc gagctcgagc 1680
301 ccgtctccgt cgccgagccc gagtccgagc ccgagtagct cgccgtcgcc gtctccgacg 1740
302 tcgtcgccgg tgtcgggtgg gctgaagggt cagtacaaga acaatgattc ggcgccgggt 1800
303 gataaccaga tcaaaccggg tctccagttg gtgaataccg ggtcgtcgtc ggtggatttg 1860
304 tcgacgggtg cgggtgcggt ctggttcacc cgggatgggt ggtcgtcgac actggtgtac 1920
305 aactgtgact gggcgccgat ggggtgtggg aatatccgcg cctcgttcgg ctcggtgaac 1980
306 ccggcgacgc cgacggcgga cacctacct gatttgaata cgggtgaata agagtgactg gtcgaatttc 2100
307 gctggtgggt cgacgggtga gattcaaac cgggtgaata agagtgactg gtcgaatttc 2100
308 accgagacca atgactactc gtatgggacg aacaccacct tccaggactg gacgaagggt 2160
309 accgtgtacg tcaacggcgt gttggtgtgg gggactgaac cgtccggcac cagccccagc 2220
310 cccacaccat ccccgagccc gagcccagc ccgagcccgg gtggggatgt gacgccgcg 2280
311 agtgtgccga ccggttgggt ggtgacgggg gtgagtgggt cgtcgggtgc gttggcgtgg 2340
312 aatgcgtcga cggataacgt ggggggtggc cattacaacg tgtaccgcaa cggggtgttg 2400
313 gtgggccagc cgacggtgac ctcgttcacc gacacgggtt tggccgcggg aaccgcgtac 2460
314 acctacacgg tggccgcggt ggacgtcgcg ggtaacacct ccgccccatc ccccccgtc 2520
315 accgccacca ccaagagtc cagccccagc cccacgccga cggggaccac ggtcaccgac 2580
316 tgcacgccc gtccctaacca gaatggtgtg accagcgtgc agggcgacga ataccgggtg 2640
317 cagaccaatg agtggaattc gtcggcccag cagtgcctca ccatcaatac cgcgaccggt 2700
318 gcctggacgg tgagcactgc gaacttcagc ggtgggaccg gcggtgcgcc cgcgacgtat 2760
319 ccgtcgatct acaagggtcg ccaactggggc aactgcacca cgaagaacgt cgggatgccg 2820
320 attcagatca gtcagattgg ttcggctgtg acgtcgtgga gtacgacgca ggtgtcgtcg 2880
321 ggcggtatg acgtggccta cgacatttgg acgaacagta cccaacgac aaccggtcag 2940
322 ccaaacggtg ccgaaatcat gatttggctg aattcgcgtg gtgggggtga gccgttcggg 3000
323 tcgacagacg cgacgggtgt gacggtcgct ggtcacacgt ggaatgtctg gcagggtcag 3060
324 cagacctcgt ggaagattat tctctacgtc ctgacccccg gtgcgacgtc gatcagtaat 3120
325 ctggatttga aggcgatttt cgcgacgcc gcggcacgcg ggtcgtccta cactccgat 3180
326 tacctgctcg acgttgaggc cgggtttgag atctggcaag gtggtcaggg cctgggcagc 3240
327 aactcgttca gcgtctccgt gacgagcggc acgtccagcc cgacaccgag cccgagcccg 3300
328 acgccgacac cgagcccagc gccgacaccg tctccgagcc cgacccccgc gccgagtcgg 3360
329 accagctcgc cgtcgtcgtc ggggtgtggc tgccgggcga cgtatgtggt gaatagtgat 3420

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/917,383

DATE: 10/01/2002

TIME: 16:15:51

Input Set : A:\4017us01.app

Output Set: N:\CRF4\10012002\I917383.raw

L:14 M:270 C: Current Application Number differs, Replaced Current Application Number